Bayesian Analysis (2015) **10**, Number 3, pp. 523–552

Generalized Quantile Treatment Effect: A Flexible Bayesian Approach Using Quantile Ratio Smoothing

Sergio Venturini*, Francesca Dominici†, and Giovanni Parmigiani‡

Abstract. We propose a new general approach for estimating the effect of a binary treatment on a continuous and potentially highly skewed response variable, the *generalized quantile treatment effect* (GQTE). The GQTE is defined as the difference between a function of the quantiles under the two treatment conditions. As such, it represents a generalization over the standard approaches typically used for estimating a treatment effect (i.e., the average treatment effect and the quantile treatment effect) because it allows the comparison of any arbitrary characteristic of the outcome's distribution under the two treatments. Following Dominici et al. (2005), we assume that a pre-specified transformation of the two quantiles is modeled as a smooth function of the percentiles. This assumption allows us to link the two quantile functions and thus to borrow information from one distribution to the other. The main theoretical contribution we provide is the analytical derivation of a closed form expression for the likelihood of the model. Exploiting this result we propose a novel Bayesian inferential methodology for the GQTE. We show some finite sample properties of our approach through a simulation study which confirms that in some cases it performs better than other nonparametric methods. As an illustration we finally apply our methodology to the 1987 National Medicare Expenditure Survey data to estimate the difference in the single hospitalization medical cost distributions between cases (i.e., subjects affected by smoking attributable diseases) and controls.

Keywords: average treatment effect (ATE), medical expenditures, National Medical Expenditures Survey (NMES), Q-Q plot, quantile function, quantile treatment effect (QTE), tailweight.

1 Introduction

The effect of a treatment on an outcome is often the main parameter of interest in many scientific fields. The standard approach used to estimate it is the so called average treatment effect (ATE), the difference between the expected values of the response's distributions under the two treatment regimes. While intuitive and useful in many situations, it suffers from some limitations; in particular, it becomes highly biased when the response is skewed.

*CERGAS, Università Bocconi, Via Röntgen 1, 20136 Milano, Italy, sergio.venturini@unibocconi.it
†Department of Biostatistics, Harvard School of Public Health, 655 Huntington Ave, Boston MA 02115, USA, fdominic@hsph.harvard.edu
‡Department of Biostatistics, Harvard School of Public Health and Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute, 44 Binney Street, Boston MA 02115, USA, gp@jimmy.harvard.edu

 DOI: 10.1214/14-BA922

A further drawback of the ATE is its coarseness as a summary of the distance between the expected value of the response's distributions under the two treatments. It is a matter of fact indeed that the effect of the treatment on the outcome often varies as we move from the lower to the upper tail of the outcome's distribution. This limitation of the ATE has been addressed in the literature by introducing the so called quantile treatment effect (QTE), the difference between the response's distribution quantiles under the two treatments (Abadie et al., 2002; Chernozhukov and Hansen, 2005; Firpo, 2007; Frölich and Melly, 2008).

In this paper we propose a more general measure of the effect of a binary treatment on a continuous outcome. We call it the *generalized quantile treatment effect* (GQTE), defined as

$$\Delta_g(p) = g(Q_1(p)) - g(Q_2(p)), \tag{1}$$

where $Q_1(p)$ and $Q_2(p)$ represent the quantile functions of the outcome under the two treatment conditions and $g(\cdot)$ is an arbitrary but known function of the quantiles. For example, if $g(\cdot)$ is chosen to be the identity function, then the GQTE simplifies to the QTE, while if $g(\cdot)$ is the integral over the percentile p, the GQTE becomes equivalent to the ATE. The GQTE is a new parameter which generalizes the existing approaches for estimating a treatment effect.

To estimate and formulate inferences about the GQTE we propose a Bayesian approach that can accommodate both symmetric and skewed outcomes, as well as situations where the sample size under a treatment condition (cases) is much smaller than the sample size under the other treatment condition (controls). In particular, we assume

$$h\left(\frac{Q_1(p)}{Q_2(p)}\right) = s(p)\,, \tag{2}$$

where h is a monotone function and s is assumed to be smooth. In other words, we assume that the transformed quantile ratio is a smooth function of the percentile p. The idea of smoothly modeling the ratio of the quantiles has been first introduced by Dominici et al. (2005), who exploited it by proposing a nonparametric estimator of the mean difference between two populations. Here we generalize their approach by permitting the comparison of any characteristic of the outcome's distributions under the two treatments.

An important theoretical contribution of this paper is the derivation of a closed form expression for the model likelihood. We show that it is possible to obtain an analytically tractable form for the Y_2 density (the controls) without explicitly specifying a model for it. Clearly the likelihood is needed to carry out the Bayesian estimation but in principle it could be employed for classic likelihood procedures as well. Moreover, our proposed approach allows one to borrow strength from one sample to the other, thus improving efficiency in the estimation of the quantiles (Dominici et al., 2005).

As an illustration, we apply our method to the comparison of the single hospitalization medical costs distribution between subjects with and without smoking attributable diseases. The data set we use is the National Medical Expenditures Survey (NMES) supplemented by the Adult Self-Administered Questionnaire Household Survey.

The paper is organized as follows. In Section 2 we define the new parameter $\Delta_g(p)$ and illustrate some quantile-based measures that will be used in the paper. In Section 3 we provide details of the estimation approach together with some special cases. We then present the results of a simulation study in Section 4 through which we conclude that under a broad set of conditions our approach performs better than other flexible methods for comparing two distributions. In Section 5 we illustrate the results of the data analysis on the NMES data set. Section 6 concludes the paper with a discussion and some final remarks.

2 The Generalized Quantile Treatment Effect (GQTE)

Consider two positive continuous random variables Y_1 and Y_2 with quantile functions Q_1 and Q_2, where

$$Q_\ell(p) \equiv F_\ell^{-1}(p) \equiv \inf\{y : F_\ell(y) \geq p\}$$

for $0 < p < 1$ and $\ell = 1, 2$. To compare F_1 and F_2 as flexibly as possible we introduce the generalized quantile treatment effect, which is defined as

$$\Delta_g(p) = g(Q_1(p)) - g(Q_2(p)), \tag{3}$$

where $g(\cdot)$ is a known function of the quantiles. Notice that no a priori assumptions are made about the admissible functions $g(\cdot)$, thus potentially any function of the quantiles can be used. Therefore, the GQTE provides a general approach to compare the response's distributions under the two treatments. More precisely, by properly choosing the function $g(\cdot)$, we can recover any specific characteristic of the outcome's distributions F_1 and F_2 and, through (3), their difference.

The simplest case arises when $g(x) = x$. In this case the GQTE simplifies to

$$\Delta(p) = Q_1(p) - Q_2(p), \tag{4}$$

the so called (unconditional) QTE (Frölich and Melly, 2008), sometimes also named the percentile-specific effect between two populations (see for example Dominici et al., 2006, 2007).

A second example is obtained by choosing $g(x) = \int x\, dp$, which produces

$$\Delta = \int_0^1 Q_1(p)\, dp - \int_0^1 Q_2(p)\, dp\,, \tag{5}$$

the extensively used ATE (see for example Wooldridge, 2010, Chapter 21).

These examples illustrate how the GQTE reduces to the two most used parameters of interest for estimating a treatment effect, the ATE and QTE. However, the GQTE can provide a variety of other useful measures. In Appendix 1 we illustrate some other interesting cases that usually are not taken into consideration in the literature.

3 Estimation Methodology

In this section we illustrate the procedure we developed for estimating the GQTE. Our proposed approach is sufficiently general that it can be used for any choice of $g(\cdot)$.

3.1 Definitions and Model Assumptions

We assume that $Y_1|\boldsymbol{\eta} \sim F_1(\cdot\,;\boldsymbol{\eta})$, where F_1 is a given probability distribution depending upon a vector of unknown parameters $\boldsymbol{\eta}$. For example, in the application presented in Section 5 we choose F_1 as a mixture distribution. To borrow information from one distribution to the other, we assume that the transformed quantile ratio is a smooth function of the percentiles with λ degrees of freedom, that is

$$h\left(\frac{Q_1(p)}{Q_2(p)}\right) = s(p,\lambda)\,, \qquad 0 < p < 1\,. \tag{6}$$

The function $h(\cdot)$ is assumed to be monotone differentiable. It represents a kind of link function and it is used to transform the quantile ratio to account for the potential skewness of the F_1 and F_2. The typical choice for skewed data is $h(x) = \log x$, while for symmetric data distributions the identity function is the most reasonable option.

For the sake of simplicity, we henceforth indicate the smooth function $s(p,\lambda)$ with reference to the corresponding design matrix $X(p,\lambda)$, so that it can be written as $X(p,\lambda)\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a vector of unknown parameters. More explicitly, we assume that $s(p,\lambda) \equiv X(p,\lambda)\boldsymbol{\beta} = \sum_{k=0}^{\lambda} X_k(p)\beta_k$, where $X_k(p)$ are orthonormal basis functions with $X_0(p) = 1$. The number of degrees of freedom λ is a further parameter that has either to be chosen or estimated from the data. In Subsection 3.7 we propose a simple approach for eliciting it. The basis functions are usually either splines or polynomials.

The main justification for assuming (6) is that it allows one to borrow information from both the response's distributions under the two treatment conditions when we estimate the GQTE $\Delta_g(p)$. Assumption (6), in fact, implies

$$Q_1(p) = Q_2(p)\, h^{-1}\left[X(p,\lambda)\,\boldsymbol{\beta}\right]\,, \tag{7}$$

and also

$$Q_2(p) = Q_1(p)\left\{h^{-1}\left[X(p,\lambda)\,\boldsymbol{\beta}\right]\right\}^{-1}\,, \tag{8}$$

which, once substituted in (3), return

$$\Delta_g(p) = g\Big(Q_2(p)\, h^{-1}\left[X(p,\lambda)\,\boldsymbol{\beta}\right]\Big) - g\Big(Q_1(p)\left\{h^{-1}\left[X(p,\lambda)\,\boldsymbol{\beta}\right]\right\}^{-1}\Big).$$

For the special case where $g(x) = x$ and $h(x) = \log(x)$, Dominici et al. (2005) have shown that under assumption (6) it is possible to obtain a more efficient estimator of Δ than the sample mean difference and the maximum likelihood estimator assuming that Y_1 and Y_2 are both log-normal.

Notice that, since the main interest in the paper resides in the estimation of $\Delta_g(p)$, for which only $\boldsymbol{\beta}$ is required, $\boldsymbol{\eta}$ is treated as nuisance (see Subsection 3.6 for further details).

In this paper we propose a Bayesian approach for estimating $\Delta_g(p)$ for any choice of $g(\cdot)$ and $h(\cdot)$. An interesting feature of our estimation procedure for $\Delta_g(p)$, is that we only need to specify the distribution function for Y_1. The specification of F_1 together with the relationship (6) automatically determines a distributional assumption for Y_2. We refer to the distribution of Y_2 induced by F_1 and assumption (6) as $F_2(\cdot\,;\boldsymbol{\beta},\boldsymbol{\eta})$.

As a last remark for this section, we want to highlight the difference between the function $g(\cdot)$, introduced in the previous section, and $h(\cdot)$, defined above in (6). They should not be confused because they have distinct roles: the former identifies the response's characteristic we want to estimate for assessing the treatment effect, while the latter has been introduced as a mechanism to attenuate the possible skewness present in the data.

3.2 Estimation Approach and Likelihood

The steps involved in our estimation approach are summarized as follows:

1. Choose a (possibly flexible) density $f_1(y_1|\boldsymbol{\eta})$ for Y_1, a smoothing function $s(p,\lambda)$ (usually a spline or a polynomial) and a value for λ;
2. From (8) derive the density function of Y_2, that we denote as $f_2(y_2|\boldsymbol{\beta},\boldsymbol{\eta})$. Note that, as proved by Theorem 1 below, this density will depend on the model parameter $\boldsymbol{\beta}$ as well as on the parameter $\boldsymbol{\eta}$ through the Y_1 density.
3. Calculate the joint likelihood $\mathbb{L}(\boldsymbol{\beta},\boldsymbol{\eta}|\boldsymbol{y}_1,\boldsymbol{y}_2)$ to use for finding the posterior distribution of $(\boldsymbol{\beta},\boldsymbol{\eta})$ in a Markov Chain Monte Carlo (MCMC) algorithm.
4. Obtain the posterior distribution of any special case of the GQTE.

The critical step in this sequence is represented by the calculation of the likelihood, which we now describe.

Consider two i.i.d. samples $(y_{11},\ldots,y_{1n_1})$ and $(y_{21},\ldots,y_{2n_2})$ drawn independently from the two populations $F_1(\cdot\,;\boldsymbol{\eta})$ and $F_2(\cdot\,;\boldsymbol{\beta},\boldsymbol{\eta})$. We refer to the former as the cases (or the treated) and to the latter as the controls (or the untreated). We assume that these distribution functions have densities $f_1(\cdot\,;\boldsymbol{\eta})$ and $f_2(\cdot\,;\boldsymbol{\beta},\boldsymbol{\eta})$ respectively. The likelihood function for our model is then given by

$$\mathbb{L}(\boldsymbol{\beta},\boldsymbol{\eta}|y_{11},\ldots,y_{1n_1},y_{21},\ldots,y_{2n_2}) = \prod_{i=1}^{n_1} f_1(y_{1i}|\boldsymbol{\eta}) \times \prod_{j=1}^{n_2} f_2(y_{2j}|\boldsymbol{\beta},\boldsymbol{\eta}). \tag{9}$$

Since we didn't state any specific distributional assumption for Y_2, in principle we could not calculate the likelihood because we don't have any expression for f_2. Two strategies are possible here. Given the f_1 specification, one possibility is to find an expression for Q_1, then map it through equation (6) to find a corresponding expression for Q_2, invert it to determine F_2, and finally differentiate the result to get f_2. Apart from simple situations, usually these steps (i.e. integration, inversion and differentiation)

need to be performed numerically. A second possibility is to replace the Y_2 density in the likelihood with its correspondent *density quantile function*, $f_2(Q_2(p_j)|\boldsymbol{\beta}, \boldsymbol{\eta})$ (see Parzen, 1979), for which the next theorem provides a closed form expression. The proof of the theorem and two additional corollaries are available in Appendix 2, while in the next subsection we provide some further explanation on how to compute f_2.

Theorem 1. *Let $Y_1|\boldsymbol{\eta} \sim F_1(\cdot\,;\boldsymbol{\eta})$, with F_1 having density function $f_1(\cdot\,;\boldsymbol{\eta})$, and assume that (6) holds. If, for every $0 < p < 1$, the vector $\boldsymbol{\beta}$ satisfies the constraint*

$$X'(p,\lambda)\,\boldsymbol{\beta}\left\{\frac{d}{d\,(X(p,\lambda)\,\boldsymbol{\beta})}h^{-1}\left[X(p,\lambda)\,\boldsymbol{\beta}\right]\right\} \leq \frac{1}{f_1\left(Q_1(p)|\boldsymbol{\eta}\right)Q_1(p)}\,, \tag{10}$$

the density quantile function $f_2(Q_2(p)|\boldsymbol{\beta}, \boldsymbol{\eta})$ for Y_2 is

$$f_2(Q_2(p)|\boldsymbol{\beta}, \boldsymbol{\eta}) = \frac{f_1\left(Q_2(p)\,h^{-1}[X(p,\lambda)\,\boldsymbol{\beta}]|\boldsymbol{\eta}\right)h^{-1}[X(p,\lambda)\,\boldsymbol{\beta}]}{1-f_1\left(Q_2(p)\,h^{-1}[X(p,\lambda)\,\boldsymbol{\beta}]|\boldsymbol{\eta}\right)X'(p,\lambda)\,\boldsymbol{\beta}\,Q_2(p)\left\{\frac{d}{d(X(p,\lambda)\,\boldsymbol{\beta})}h^{-1}[X(p,\lambda)\,\boldsymbol{\beta}]\right\}}\,. \tag{11}$$

The function $f_2(Q_2(p)|\boldsymbol{\beta}, \boldsymbol{\eta})$ is a properly defined density.

Note that f_2 correctly depends upon both the model parameter $\boldsymbol{\beta}$ and the Y_1 parameter $\boldsymbol{\eta}$ through the f_1 density. The motivation for the constraint (10) comes from the need to guarantee that f_2 is a non-negative function. As a further remark, we observe that the term in the likelihood involving $f_2(Q_2(p_j)|\boldsymbol{\beta}, \boldsymbol{\eta})$ depends upon the observations y_{2j} through the unknown quantile function values $Q_2(p_j)$.

3.3 Details for the Computation of f_2

A computational drawback of our proposal is that the "true" values of the percentiles p_j, i.e. those generated under the assumed model for Y_2, should be used in the calculation of the likelihood. Unfortunately, these are not available, because the cumulative distribution function F_2 is not given explicitly and we cannot find the p_j corresponding to the observed data y_{2j} as $F_2(y_{2j}) = p_j$.

The approach we recommend to bypass this issue is to approximate the p_j using the procedure described in Gilchrist (2000), which we summarize as follows:

1. Denoting with $y_{2(j)}$ the ordered observed values for Y_2, we look for the corresponding set of ordered $p_{(j)}$ such that $y_{2(j)} = \widehat{Q}_2(p_{(j)})$, where $\widehat{Q}_2(p)$ is an estimate of $Q_2(p)$ based on the current values of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ (i.e. the values from the current MCMC draw). More specifically, we find the $p_{(j)}$ using the following procedure: suppose p_0 is the current estimate of p for a given y value. Then, for a value of p close to p_0, $\widehat{Q}_2(p)$ can be approximated using the following Taylor series expansion

$$\begin{aligned}\widehat{Q}_2(p) &= \widehat{Q}_2(p_0) + \widehat{Q}_2'(p_0)(p-p_0)\\ &= \widehat{Q}_2(p_0) + \widehat{q}_2(p_0)(p-p_0),\end{aligned}$$

which, solving for p, gives

$$p = p_0 + \frac{y - \widehat{Q}_2(p_0)}{\widehat{q}_2(p_0)}, \tag{12}$$

where $\widehat{q}_2(p_0)$ is the quantile density function corresponding to $\widehat{Q}_2(p_0)$ and where we used the fact that $y = \widehat{Q}_2(p)$. As a starting point for $p_{(j)}$ we use $j/(n_2+1)$, $j = 1, \ldots, n_2$. Equation (12) is used in an iterative fashion till the given value of $\widehat{Q}_2(p)$ differs from y by less than some chosen small amount (we use 10^{-8}).

2. Once the values of p_j are available, we compute the quantities $f_2(Q_2(p_j)|\boldsymbol{\beta}, \boldsymbol{\eta})$ using equation (11). The critical issue in this step is the calculation of the derivative $X'(p, \lambda)$. In the cases we consider here (i.e. either a polynomial or a spline basis), the derivative is available in closed form and so no further numerical approximation is needed.

Strictly speaking, the calculation of f_2 provided by the procedure we just described is not exact but involves a numerical approximation. We performed a detailed analysis on the goodness of this approximation and we found that the actual and approximated f_2 values (and hence the overall likelihood) were indistinguishable.

3.4 Special Cases

We present now some special cases where an appropriate choice of the design matrix $X(p, \lambda)$ allows to recover an exact expression for $f_2(Q_2(p)|\boldsymbol{\beta}, \boldsymbol{\eta})$ belonging to a known distribution family. The proofs of these special cases are provided in Appendix 3.

Case 1: Y_1 is Uniform and $X(p, \lambda = 0) = 1$. In this case we assume that $Y_1|\theta_1 \sim \mathcal{U}[\,0, \theta_1]$ and choose $h(x) = x$. Then $Q_1(p)/Q_2(p) = \beta_0$ and from (29) it follows that

$$f_2(Q_2(p)|\theta_1, \beta_0) = \frac{\beta_0}{\theta_1} \mathbb{I}_{[0,\, \theta_1/\beta_0]}\{Q_2(p)\}\,,$$

which corresponds to the density quantile function of a uniform random variable with parameter $\theta_2 = \theta_1/\beta_0$, where $\mathbb{I}_A\{x\}$ denotes the indicator function taking value 1 if $x \in A$ and 0 otherwise. Note that it correctly depends both upon the parameter β_0 and the Y_1 parameter $\eta = \theta_1$.

Case 2: Y_1 is Log-normal and $X(p, \lambda = 1) = [1, \Phi^{-1}(p)]$. We now assume $Y_1|\mu_1, \sigma_1^2 \sim \mathcal{L}n(\mu_1, \sigma_1^2)$ and fix $h(x) = \log(x)$. It follows that $\log\{Q_1(p)/Q_2(p)\} = \beta_0 + \beta_1\, \Phi^{-1}(p)$, where $\Phi^{-1}(p)$ is the quantile function of a standard normal random variable. Then by (31) we get

$$f_2(Q_2(p)|\mu_1, \sigma_1^2, \beta_0, \beta_1) = \frac{1}{Q_2(p)\sqrt{2\pi}(\sigma_1 - \beta_1)} \exp\left\{ -\frac{[\log Q_2(p) - (\mu_1 - \beta_0)]^2}{2(\sigma_1 - \beta_1)^2} \right\},$$

which is the density quantile function of a $\mathcal{L}n(\mu_2, \sigma_2^2)$ random variable with $\mu_2 = (\mu_1 - \beta_0)$ and $\sigma_2 = (\sigma_1 - \beta_1)$. Note that the density quantile function of Y_2 correctly depends

both upon the parameters $\boldsymbol{\beta} = (\beta_0, \beta_1)$ and the Y_1 parameters $\boldsymbol{\eta} = (\mu_1, \sigma_1^2)$. In this case the constraint (30) simply requires that $\beta_1 \leq \sigma_1$, for every $0 < p < 1$.

Case 3: Y_1 is Pareto and $X(p, \lambda = 1) = [1, \log(1-p)]$. Suppose $Y_1 | a_1, b_1 \sim \mathcal{P}a(a_1, b_1)$ and choose $h(x) = \log(x)$. In this case $\log\{Q_1(p)/Q_2(p)\} = \beta_0 + \beta_1 \log(1-p)$ and by (31) we get

$$f_2(Q_2(p)|a_1, b_1, \beta_0, \beta_1) = \frac{a_1}{a_1\beta_1 + 1} \left(b_1 e^{-\beta_0}\right)^{\frac{a_1}{a_1\beta_1+1}} Q_2(p)^{-\left(\frac{a_1}{a_1\beta_1+1}+1\right)},$$

which represents the density quantile function of a $\mathcal{P}a(a_2, b_2)$ random variable with $a_2 = \frac{a_1}{a_1\beta_1+1}$ and $b_2 = b_1 e^{-\beta_0}$. The density quantile function of Y_2 correctly depends both upon the parameters $\boldsymbol{\beta} = (\beta_0, \beta_1)$ and the Y_1 parameters $\boldsymbol{\eta} = (a_1, b_1)$. In this case the constraint (30) requires that $\beta_1 \geq -\frac{1}{a_1}$, for every $0 < p < 1$.

3.5 Prior Structure and Posterior Calculation

The parameters $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ are assumed to be a priori independent, that is

$$p(\boldsymbol{\beta}, \boldsymbol{\eta} | \zeta_{\boldsymbol{\beta}}, \zeta_{\boldsymbol{\eta}}) = p(\boldsymbol{\beta}|\zeta_{\boldsymbol{\beta}}) \times p(\boldsymbol{\eta}|\zeta_{\boldsymbol{\eta}}),$$

where $\zeta_{\boldsymbol{\beta}}$ and $\zeta_{\boldsymbol{\eta}}$ are the prior hyperparameters for $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ respectively. For $p(\boldsymbol{\beta}|\zeta_{\boldsymbol{\beta}})$ we use a a multivariate normal distribution with mean equal to the ordinary least squares (OLS) estimate of $\boldsymbol{\beta}$ based on the model

$$h\left(\frac{y_{1(i)}}{y_{2(i)}}\right) = X(p_i, \lambda)\boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{13}$$

where $n = \min(n_1, n_2)$, $p_i = i/(n+1)$ and variance-covariance matrix equal to $\sigma^2_{\boldsymbol{\beta}} I_{(\lambda+1)}$, where $\sigma^2_{\boldsymbol{\beta}}$ is normally fixed at a high value to induce a weakly informative prior distribution for each β_j and $I_{(\lambda+1)}$ indicates the identity matrix with size $(\lambda+1)$. For the prior distribution for $\boldsymbol{\eta}$, the choice clearly depends upon the assumption made about F_1, but we suggest to use conjugate priors. For an example see the application in Section 5.

The posterior distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ are obtained by an MCMC simulation. In particular, we use an independent Metropolis-Hastings algorithm with blocking over $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ separately (see Gilks et al., 1996; Robert and Casella, 2004; O'Hagan and Forster, 2004; Carlin and Louis, 2009). As the proposal distribution for $\boldsymbol{\beta}$ we use a $(\lambda+1)$-dimensional t distribution with mean and scale matrix chosen to match the $\boldsymbol{\beta}$ OLS estimate and variance from (13), and a small number of degrees of freedom, usually set to 3. As with the prior, the proposal distribution for $\boldsymbol{\eta}$ depends upon the particular application under investigation (see Section 5 for an example).

3.6 GQTE Estimation and Inference

Once the $\boldsymbol{\beta}$ parameters have been estimated and the convergence of the simulated chains has been assessed by conventional methods (see Carlin and Louis, 2009, or Gelman et al.,

2013), we can obtain the posterior distribution of the GQTE for any choice of $g(\cdot)$ as we now describe.

For each iteration m of the MCMC simulation, a value $\widehat{\boldsymbol{\beta}}^{(m)}$ for $\boldsymbol{\beta}$ is available. Using expressions (7) and (8) we can obtain $\widehat{Q}_1^{(m)}(p)$ and $\widehat{Q}_2^{(m)}(p)$ as

$$\widehat{Q}_1^{(m)}(p_{2i}) = y_{2(i)} h^{-1}\left[X\left(p_{2i},\lambda\right)\widehat{\boldsymbol{\beta}}^{(m)}\right], \qquad i = 1,\ldots,n_2,$$

and

$$\widehat{Q}_2^{(m)}(p_{1i}) = y_{1(i)}\left\{h^{-1}\left[X\left(p_{1i},\lambda\right)\widehat{\boldsymbol{\beta}}^{(m)}\right]\right\}^{-1}, \qquad i = 1,\ldots,n_1,$$

where the $y_{\ell(i)}$ are the order statistics for sample ℓ, while $p_{\ell i} = i/(n_\ell + 1)$, $i = 1,\ldots,n_\ell$, for $\ell \in \{1,2\}$. It then follows that the m-th iteration value for $\Delta_g(p)$ is given by

$$\widehat{\Delta}_g^{(m)}(p) = g\left(\widehat{Q}_1^{(m)}(p)\right) - g\left(\widehat{Q}_2^{(m)}(p)\right), \qquad 0 < p < 1, \tag{14}$$

where $\widehat{Q}_1^{(m)}(p)$ and $\widehat{Q}_2^{(m)}(p)$ are found by interpolating the estimated quantile functions $\left(p_{2i}, \widehat{Q}_1^{(m)}(p_{2i})\right)$ and $\left(p_{1i}, \widehat{Q}_2^{(m)}(p_{1i})\right)$. The estimate of $\Delta_g(p)$ is finally obtained through the Rao-Blackwellized estimator

$$\widehat{\Delta}_g(p) = \frac{1}{M}\sum_{m=1}^{M} \widehat{\Delta}_g^{(m)}(p), \qquad 0 < p < 1, \tag{15}$$

where M is the total number of iterations. Since the whole posterior distribution of $\Delta_g(p)$ is available, standard inferential questions can be easily addressed in the usual ways.

As detailed in Section 2, many interesting special cases arise from the general definition of the GQTE. For example, if interest lies in estimating the QTE defined in (4), $g(\cdot)$ corresponds to the identity function and expression (14) becomes

$$\widehat{\Delta}^{(m)}(p) = \widehat{Q}_1^{(m)}(p) - \widehat{Q}_2^{(m)}(p), \tag{16}$$

which can be evaluated for any value of $p \in (0,1)$.

If the focus is on the ATE, defined in (5), then (14) returns the estimator

$$\widehat{\Delta}^{(m)} = \frac{1}{n_2}\sum_{i=1}^{n_2} y_{2(i)} h^{-1}\left[X\left(p_{2i},\lambda\right)\widehat{\boldsymbol{\beta}}^{(m)}\right] - \frac{1}{n_1}\sum_{i=1}^{n_1} y_{1(i)}\left\{h^{-1}\left[X\left(p_{1i},\lambda\right)\widehat{\boldsymbol{\beta}}^{(m)}\right]\right\}^{-1}. \tag{17}$$

Appendix 4 contains the details for some other cases. As a final remark, note that an appealing feature of this approach is that the MCMC procedure needs to be run only once to compute the difference between any measure of the treatment effect of interest in the two groups.

3.7 Selecting the Number of Degrees of Freedom λ

The choice of the number of degrees of freedom λ to use in the procedure above is not trivial. Many approaches can be proposed, but to keep the computational burden manageable, we propose to elicit it by minimizing an empirical version of the L_1 discrepancy measure (Devroye and Lugosi, 2001)

$$D^0(\lambda) = \sum_{i=1}^{n} \left| f_2(Q_2(p_i)|\boldsymbol{\beta}, \boldsymbol{\eta}) - f_2^0(Q_2(p_i)) \right|, \tag{18}$$

where f_2^0 denotes the unknown true Y_2 density. More precisely, we select λ using the following procedure: for each $\lambda \in \{1, \ldots, \lambda_{\max}\}$, where $\lambda_{\max}$ is the maximum admissible value for λ, we estimate $f_2(Q_2(p_i)|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\eta}})$, where $\widehat{\boldsymbol{\beta}}$ is equal to the OLS estimate given in (13) and $\widehat{\boldsymbol{\eta}}$ is estimated by using only the data $(y_{11}, \ldots, y_{1n_1})$. After replacing the true unknown density f_2^0 with a kernel density estimate of the data $(y_{21}, \ldots, y_{2n_2})$, the value of λ is chosen as that minimizing the value of (18) over the set $\{1, \ldots, \lambda_{\max}\}$. One drawback of this approach is that it tends to select high values of λ. We provide further discussion about this issue in Section 6.

4 Simulation Study

In this section we report the results of a simulation study we performed to compare the finite sample properties of our method with those of other flexible approaches.The simulation indicates that often the GQTE procedure has lower mean squared error and a similar bias as other flexible methods for comparing two distributions. In particular, we contrast our proposal with the smooth quantile ratio estimation (SQUARE) approach presented in Dominici et al. (2005), and the Probit stick-breaking process (PSBP) proposed in Chen and Dunson (2009). The former is a frequentist semiparametric method while the latter is a Bayesian nonparametric model. We now provide some details about these two methodologies and a justification for using them.

In SQUARE it is assumed that the log quantile ratio is a smooth function of the percentile p with λ degrees of freedom, that is

$$\log\left\{\frac{Q_1(p)}{Q_2(p)}\right\} = s(p, \gamma), \qquad 0 < p < 1.$$

The basic idea of smooth quantile ratio estimation is to replace the empirical quantiles with smoother versions obtained by smoothing the log-transformed ratio of the two quantile functions across percentiles. SQUARE has been proposed by Dominici et al. (2005) as an estimator of the mean difference between two populations with the advantage of providing substantially lower mean squared error and bias than the sample mean difference or the maximum likelihood estimator for log-normal populations. To estimate γ a B-fold cross-validation approach is suggested. Finite sample inference is performed by bootstrap but they also provide large sample results.

The PSBP is a general nonparametric Bayesian model which has been proposed by Chen and Dunson (2009) for estimating the conditional distribution of a response

variable given multiple predictors. More specifically, the PSBP is a prior for an uncountable collection of random distributions. Like for the models belonging to the class of dependent Dirichlet processes (MacEachern, 1999), the PSBP main idea is to allow for dependence across a family of related distributions as a function of some covariates. More explicitly, the PSBP induce dependence in the weights of the stick-breaking representation (Sethuraman, 1994) by replacing the beta-distributed random variables with a probit model. This simple change greatly enhances the flexibility of the model, thus providing an extremely interesting extension within the framework of dependent priors across families of probabilities measures.

We decided to compare the GQTE with these two methods because they are both highly flexible and have been proved to perform well under a broad set of situations. Originally the simulation also included the ANOVA dependent Dirichlet process mixtures proposed by De Iorio et al. (2004), but we decided not to report it here because it performed poorly as compared to the PSBP model. The reason for such inferior results resides in the definition of the model itself, which assumes the weights in the stick-breaking representation of the process to be fixed, i.e. the same for the response distributions under the two treatment conditions.

Since SQUARE produces an estimate only for the mean difference between two populations, our simulation is restricted to this specific case. We are aware that the results only provide a partial demonstration of the GQTE advantages over the other methods, but we also need to stress that the ATE is the most common measure used in practice for estimating the extent of a treatment effect.

Our simulation framework is similar to that used in Dominici et al. (2005) and includes five scenarios, which are described in Table 1 under the labels A to E. In scenarios A, B and C the Y_2 distribution is assumed to be log-normal with parameters $\mu_2 = 7$ and $\sigma_2 = 1.5$ which approximately correspond to the sample statistics for the medical expenditures of non-diseased subjects from the NMES data set. In scenario A, the Y_1 distribution is also log-normal but with larger values of the parameters, namely $\mu_1 = 7.5$ and $\sigma_1 = 1.75$. Scenarios B and C use a different assumption for the Y_1 distribution chosen to represent some reasonable shapes. The next two scenarios, D and E, compare the performances of the different methods using real data. In particular, in scenario D the data are randomly drawn from the distributions of nonzero Medicare expenditures for cases and controls from the NMES data set. Finally, scenario E assumes that both populations follow a gamma distribution with finite second moment.

Under each scenario we compare the mean squared error (RMSE) and bias (RB) in percentage relative to the sample mean difference $(\bar{y}_1 - \bar{y}_2)$ for the following methods: (1) the GQTE approach that assumes Y_1 to be log-normally distributed, (2) the GQTE assuming Y_1 follows a gamma distribution, (3) the SQUARE method using natural cubic splines with the number of degrees of freedom chosen by 10−fold cross-validation, (4) the PSBP model using the treatment indicator as the only predictor. The GQTE estimators use a natural cubic spline basis for the cubic-root transformed quantile-ratio smoother with the number of degrees of freedom λ chosen following the procedure detailed in Section 3.7. The RMSE is computed by $[\{\mathrm{MSE}(\bar{y}_1-\bar{y}_2)-\mathrm{MSE}(\widehat{\Delta})\}/\mathrm{MSE}(\bar{y}_1-\bar{y}_2)]\times 100$,

Table 1: Simulation Study – Sampling mechanisms under each simulation scenario. In scenario D, $\widehat{F}_g$ $(g = 1, 2)$ are the empirical cumulative distribution functions of the nonzero medical expenditures for patients in the case and control groups from the NMES data set, and, in scenarios B and C, $g(u) = exp\{7 + 1.5 \cdot \Phi^{-1}(u)\}$. Moreover, in scenario B, $s_B(u) = \mathbb{I}_{(0,1)}(u) + \mathbb{I}_{(0.9,1)}(u)$, while in scenario C, $s_C(u) = 8u(1-u)\mathbb{I}_{(0,1)}(u)$.

Scenario	Population 1	Population 2	n_1	n_2
A	$\mathcal{L}og\mathcal{N}(7.5, 1.75)$	$\mathcal{L}og\mathcal{N}(7, 1.5)$	100	1000
B	$u \sim \mathcal{U}nif(0,1)$, $y_1 = g(u)e^{s_B(u)}$	$\mathcal{L}og\mathcal{N}(7, 1.5)$	100	1000
C	$u \sim \mathcal{U}nif(0,1)$, $y_1 = g(u)e^{s_C(u)}$	$\mathcal{L}og\mathcal{N}(7, 1.5)$	100	1000
D	$\widehat{F}_1$	$\widehat{F}_2$	100	1000
E	$\mathcal{G}a(2.5, 2.5/\bar{y}_1)$	$\mathcal{G}a(2.5, 2.5/\bar{y}_2)$	100	1000

while the RB is defined as $[\{\mathbb{E}(\widehat{\Delta}) - \Delta\}/\Delta] \times 100$. Note that positive values for the RMSE imply a better performance for the estimator as compared to the sample mean difference.

The results for 100 generated data sets for each scenario are reported in Table 2. We considered only the case of unbalanced samples with $n_1 = 100$ and $n_2 = 1000$ because typically it represents a more critical situation to deal with in practice. These results show that the GQTE has a smaller mean squared error in most of the scenarios considered.

Table 2: Simulation Study – Results from 100 replicate datasets. RMSE is the mean squared error relative to $(\bar{y}_1 - \bar{y}_2)$ in percentage defined by $[\{\text{MSE}(\bar{y}_1 - \bar{y}_2) - \text{MSE}(\widehat{\Delta})\}/\text{MSE}(\bar{y}_1 - \bar{y}_2)] \times 100$, and RB is the bias relative to $(\bar{y}_1 - \bar{y}_2)$ in percentage defined by $[\{\mathbb{E}(\widehat{\Delta}) - \Delta\}/\Delta] \times 100$, under the data generation mechanisms described in Table 1. The splines degrees of freedom λ for the GQTE approach are chosen using the heuristic algorithm described in Section 3.7 while for SQUARE we use 10-fold cross-validation.

	Scenario A		Scenario B		Scenario C		Scenario D		Scenario E	
	RMSE	RB	RMSE	RB	RMSE	RB	RMSE	RB	RMSE	RB
GQTE ($\mathcal{L}og\mathcal{N}$)	39	-29	-26	-23	3	13	-26	18	-1	-2
GQTE ($\mathcal{G}amma$)	33	-16	1	-9	48	7	25	0	0	-2
SQUARE	35	-6	-7	-5	1	13	25	3	0	-1
PSBP	1	-6	6	-10	-8	9	-13	6	-3	-3
$\text{MSE}(\bar{y}_1 - \bar{y}_2)$	2952		5992		1051		2100		753	
Δ	4982		15225		5244		7144		7144	

In scenario A, where both the populations are log-normal, the GQTE assuming Y_1 is log-normally distributed performs around 40% better than $(\bar{y}_1 - \bar{y}_2)$, slightly better than SQUARE, even if somewhat biased. This result is superior to that of PSBP, which performs approximately as well as the sample mean difference. In scenario B, the PSBP provides the best result with a mean square error which is 6% smaller than $(\bar{y}_1 - \bar{y}_2)$, followed by the GQTE with gamma distributed Y_1. In scenarios C, D and E the GQTE with gamma distributed Y_1 outperforms both the PSBP and SQUARE. More

specifically, in scenario C the GQTE provides a mean square error that is approximately 50% smaller than $(\bar{y}_1 - \bar{y}_2)$. This is also the least biased result. In scenarios D and E, the GQTE approach provides comparable results as those provided by SQUARE.

5 Application: Medical Costs for Smoking Attributable Diseases

As an illustration, we apply the GQTE approach to the NMES data, where the distributions of Y_1 (the cases) and Y_2 (the controls) are highly right-skewed. For this reason, we decide to use $h(x) = \log(x)$. We show that having a smoking attributable disease induces both a location and scale shift in the medical expenditure distribution as compared to that for non-affected subjects, but with a thinning of the corresponding distribution's tails.

5.1 Data Description

The data used in the following analysis is taken from the National Medical Expenditure Survey (NMES) and have been previously studied by other authors (for example Dominici et al., 2005). It provides data on annual medical expenditures, disease status, age, race, socio-economic factors, and critical information on health risk behaviors such as smoking, for a representative sample of U.S. non-institutionalized adults (National Center For Health Services Research, 1987). NMES data derive from the 1987 wave. In the data set used here a total of 9,416 individuals are available. Table 3 briefly summarizes the data set (numbers in parentheses represent the percentage of subjects with non-zero expenditures).

Table 3: Disease cases and controls for smokers (current or former) and for non-smokers. Numbers within parentheses represent the percentage of people in that cell with non-zero expenditures.

	smokers	non smokers	Total
cases	165 (62%)	23 (70%)	188 (63%)
controls	4,682 (21%)	4,546 (28%)	9,228 (25%)
Total	4,847 (22%)	4,569 (28%)	9,416 (25%)

We consider as cases (Y_1) those individuals who are affected by smoking diseases, namely lung cancer and chronic obstructive pulmonary disease, while the controls (Y_2) are persons without a major smoking attributable disease.

In the following analyses we consider only the non-zero costs paid for each hospitalization by diseased and non-diseased subjects.

Figures 1(a) and (c) show the histograms and boxplots for the medical costs of the cases and controls. Both the distributions are highly right-skewed, with the cases sample which is much smaller than the controls one (118 vs. 2,262). Table 4 contains

some high-order sample quantiles for the two groups which confirm the heavier tails of the cases costs distribution. However, note also that the controls sample has a higher maximum cost.

Table 4: Summary of the NMES data set: high-order quantiles of non-zero medical expenditures for cases and controls.

Quantile order	75	90	95	99	99.9	100
Quantile for cases ($)	11,525.17	29,439.96	49,595.77	63,886.05	213,567.69	233,047.63
Quantile for controls ($)	2,600.00	9,799.664	30,625.206	49,771.60	135,896.07	238,185.94

Figures 1(b) and (d) show the histograms and boxplots for the cubic root transformed data. The need for such a transformation derives from the particular choice we make regarding the cases distribution (see next subsection) and is not a general requirement of our approach. Moreover, the use of this transformation does not alter in any way the results and the conclusions we draw, hence in the following we systematically refer to the transformed data. At any rate, note also that, even after the transformation, the outcome distributions still present heavy right tails. This conclusion motivates the use of $h(x) = \log(x)$.

In Figure 2(a) we report the Q-Q plot for the (cubic root transformed) NMES data. We can identify a non-linear smooth relationship between the cases and controls medical expenditures. Panel (b) of the same picture, which shows the quantile ratio as a function of the percentile p, confirms these findings.

5.2 Model Assumptions and Tuning Parameters

In this application we assume that $Y_1|\boldsymbol{\pi}, \theta \sim \mathcal{GSM}(\boldsymbol{\pi}, \theta|J)$, a particular mixture of gamma distributions with density

$$f(y|\pi_1, \ldots, \pi_J, \theta) = \sum_{j=1}^{J} \pi_j \frac{\theta^j}{\Gamma(j)} y^{j-1} e^{-\theta y},$$

where the mixing occurs over the shape parameters and where J, the number of components, is fixed a priori. First introduced in Venturini et al. (2008), it has been explicitly developed as a model for right-skewed distributions and its parameterization allows to create a convenient and flexible method characterized by a single scale parameter for all the gamma components, plus the ordinary set of mixture weights. We use conjugate priors $\theta \sim \mathcal{G}a(\alpha, \delta)$ and $\boldsymbol{\pi} \sim \mathcal{D}_J\left(\frac{1}{J}, \ldots, \frac{1}{J}\right)$ for the shared scale parameter and the mixture weights respectively. The number of mixture components is fixed at $J = 40$, while the θ hyperparameters are set to $\alpha = 845$ and $\delta = 1,300$ (for more information on the elicitation of these priors see Venturini et al., 2008, Section 2.3).

The initial values of the Metropolis-Hastings algorithm are chosen as follows: the $\boldsymbol{\beta}$ chain is started from its OLS estimate, as discussed in (13), while for $\boldsymbol{\eta} = (\theta, \boldsymbol{\pi})$ we first get a preliminary estimate (with $5,000$ iterations) using the approach described

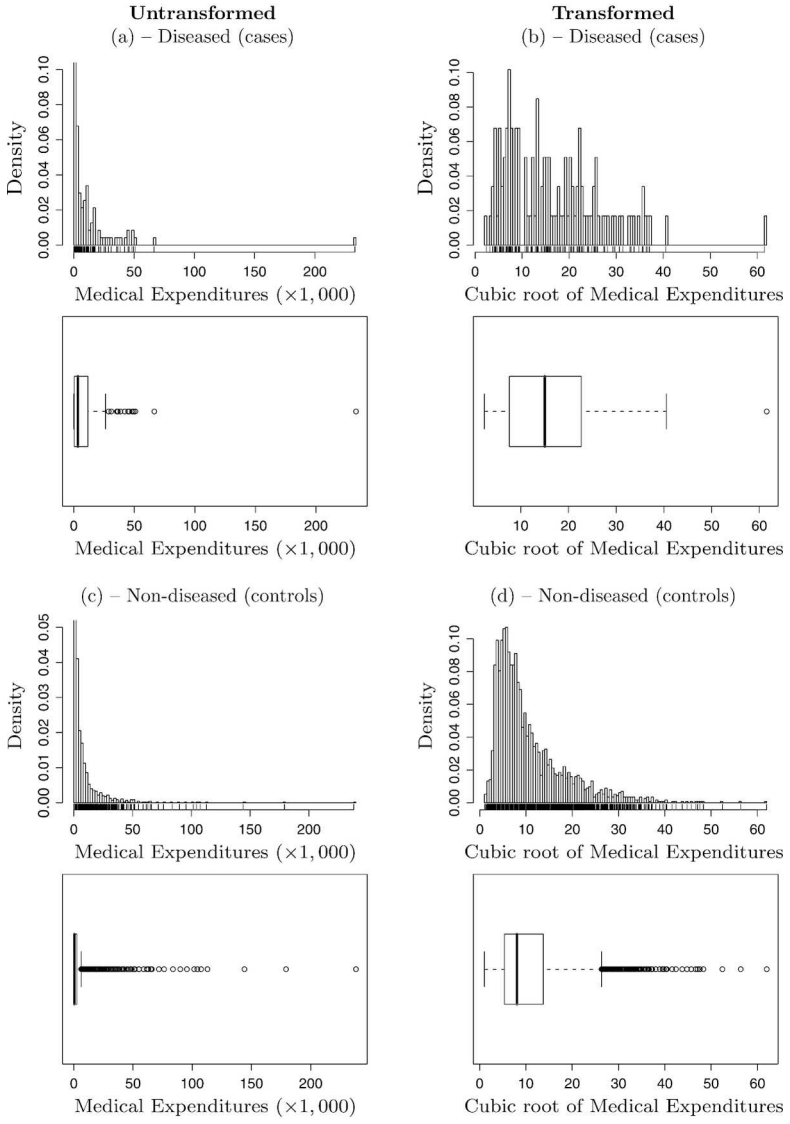

Figure 1: Histograms and boxplots of positive medical expenditures for hospitalizations regarding smoking attributable diseases (lung cancer and coronary obstructive pulmonary disease) from the 1987 National Medicare Expenditure Survey (for clarity of exposition, the histogram of the original expenditures has been truncated at the top).

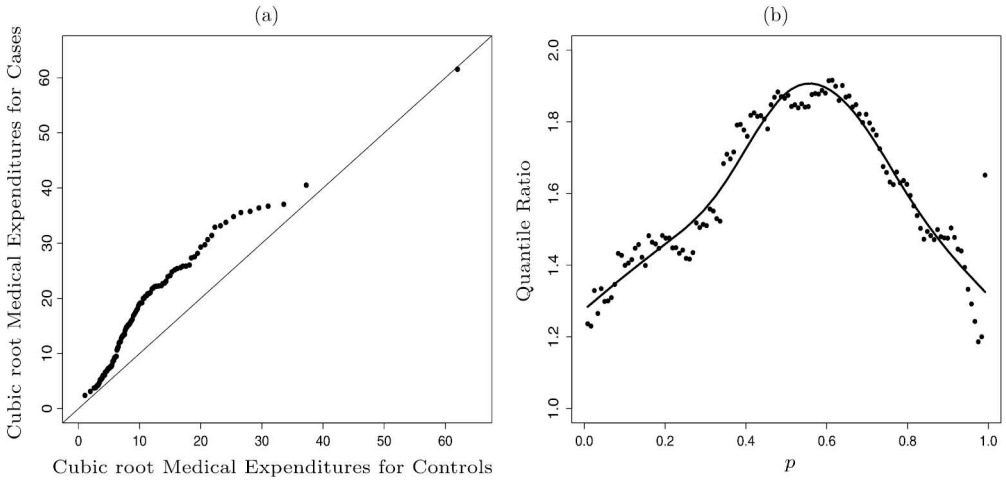

Figure 2: (a) Q-Q plot of cubic root transformed non-zero medical expenditures. (b) Quantile ratio across percentiles with a fitted natural cubic spline.

in Venturini et al. (2008), and then we fix their starting values to the corresponding estimated posterior averages.

We run the MCMC algorithm for $1,000,000$ iterations plus $200,000$ iterations as burn-in. Such a large number of iterations is necessary because the model, being quite complicated, has shown a slow convergence behavior of the chains.

5.3 Results

The selection procedure described in Subsection 3.7 for the number of degrees of freedom suggested a value of λ equal to 6, which can be considered fairly satisfactory from a visual inspection of the scatterplot (see Figure 2(b)).

The acceptance rates for the MCMC posterior simulations are relatively small, being around 0.5% for $\boldsymbol{\beta}$, 25% for θ and 1.6% for $\boldsymbol{\pi}$. Despite that, we do not consider these results as problematic since the chain is moving in a high-dimensional space ($(\lambda + 1) + (J + 1) = 48$ dimensions), which necessarily slows down the convergence process. This is the main reason why we decide to run the simulation for a longer time. However, the results of the analysis presented below indicate that convergence was attained. We made other attempts with simpler (but less flexible) specifications of the Y_1 distribution, which showed a more conventional behavior of the acceptance rates.

Figure 3 shows the fitted values for the estimated model (6). The gray dots represent the quantile ratio for the transformed data as a function of the percentile p. The solid line illustrates the estimated posterior mean of the quantile ratio, while the dashed one represents its OLS estimate (the same line as in Figure 2(b)). The shaded

area gives the credible bands for the estimated posterior means, showing a fairly low amount of uncertainty around the estimates. Moreover, from the picture we can see that our model is less sensitive to extreme observations, especially in the right tails of the distributions.

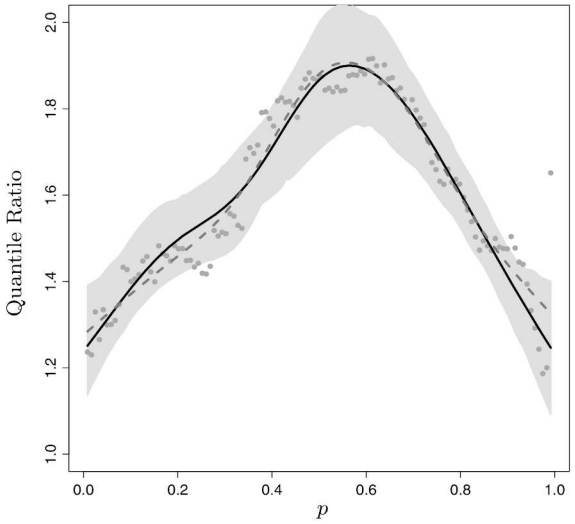

Figure 3: Fitted values of the estimated model (6). The solid line represents the estimated pointwise posterior means, while the shaded area corresponds to their pointwise 95% credible intervals. The dashed line corresponds to the OLS fit for the same data, as described in (13).

In Figure 4 we report the estimated Y_2 density. It is possible to ascertain a quite good fit. In the display, together with the $f_2(Q_2(p_{2i})|\theta, \boldsymbol{\pi}, \boldsymbol{\beta})$ posterior mean, we put the corresponding 95% credible bands and the histogram of the data.

We now describe the results for the GQTE $\Delta_g(p)$ introduced in Section 3 for some choices of the function $g(\cdot)$. We start from the QTE, denoted as $\widehat{\Delta}(p)$ in (16), whose estimate is shown in Figure 5. The solid line represents the posterior mean of the medical costs QTE between cases and controls and the gray area is the corresponding 95% credible interval, while the dashed line portrays the sample quantile differences. We can see that the distribution of the medical expenditures for subjects with smoking attributable diseases is always above that of those without smoking-related diseases. However, a much larger variability results in estimating the difference for the very extreme quantiles. This behavior is not too surprising since the two samples become very sparse as the medical expenditures become bigger (see the boxplots in Figure 1).

Figures 6(a) and (b) contain the posterior distributions of the ATE and the standard deviation difference, as defined in Section 2. The sample mean and standard deviation differences are depicted in the two plots with a vertical dotted line, while the 95% credible intervals are indicated using dashed lines. The ATE estimated posterior mean

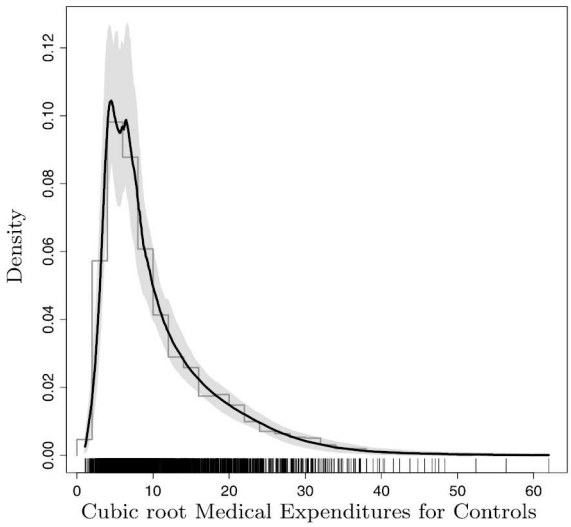

Figure 4: Estimated Y_2 density. The solid line represents the estimated pointwise posterior means, while the shaded area shows the corresponding 95% credible intervals. The thinner dark gray line depicts the data histogram.

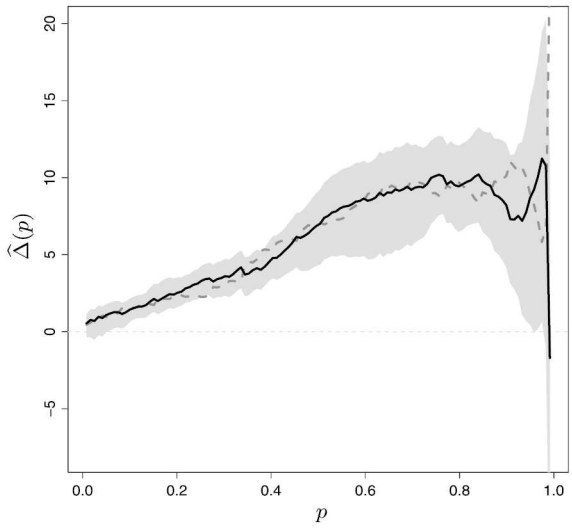

Figure 5: Estimated Quantile Treatment Effect (QTE), defined as $\Delta(p) = Q_1(p) - Q_2(p)$. The solid line reports the estimated pointwise posterior means, the shaded area gives the corresponding 95% credible intervals, while the dashed line shows the sample quantile differences. Data are cubic-root transformed.

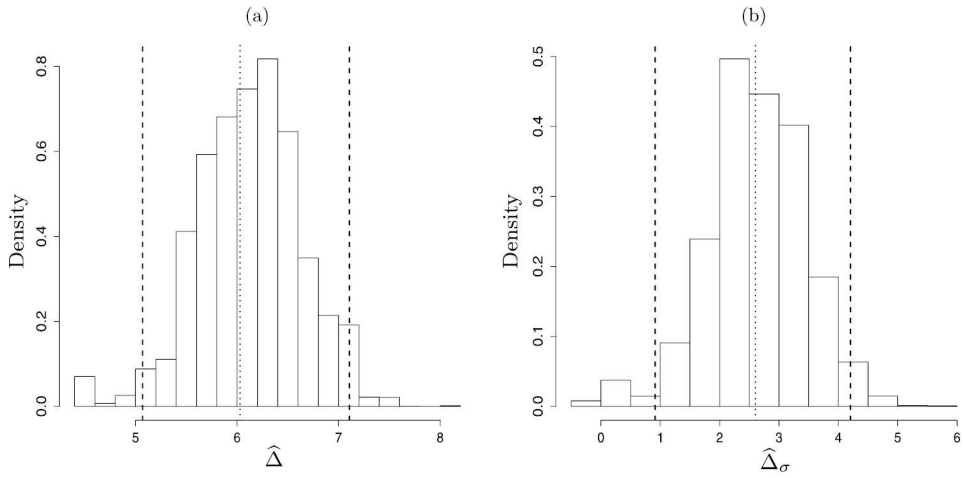

Figure 6: (a) Estimated posterior distribution of the Average Treatment Effect (ATE) between cases and controls medical expenditures (cubic root transformed), as defined in (5). The vertical dashed lines represent the 95% credible interval, while the dotted line is the sample mean difference. (b) Estimated posterior distribution of the standard deviation difference between cases and controls medical expenditures (cubic root transformed), as defined in Section 2. The vertical dashed lines represent the 95% credible interval, while the dotted line depicts the sample standard deviation difference.

(on the log scale) is equal to 6.1127, while the estimated posterior mean of the standard deviation difference is 2.6275. These results prove that having a smoking attributable disease has a significant negative impact on both the location and scale of the single hospitalization medical cost distribution. After re-transforming the estimated quantiles on the original scale, we get an estimated posterior mean for the ATE between diseased and non-diseased subjects equal to $\$6,244.10$.

Finally, Figure 7 shows the impact of the treatment variable (i.e., having or not a smoking attributable disease) on the tailweight functions $TW(p)$, defined in (22), of the two populations. The tails of the medical costs distribution for the diseased subjects tend to be heavier than those of the non-diseased ones for values of p up to approximately 0.6, but the situation is inverted as we move to consider higher percentiles. Hence, while the fact of being affected by smoking attributable diseases tends to increase both the average and the variance of the medical expenditures distribution, we have found that the opposite occurs to the tail probabilities, that is, to the chances of incurring very high medical costs in a single hospitalization.

As a last comment, we would like to remark on the explicit choice we made to exclude the observations with null medical costs. We took this decision because the inclusion of this further feature of the data requires the extension of our approach to a two-part modeling framework (Mullahy, 1998; Cameron and Trivedi, 2005), which doesn't appear to be straightforward in our context.

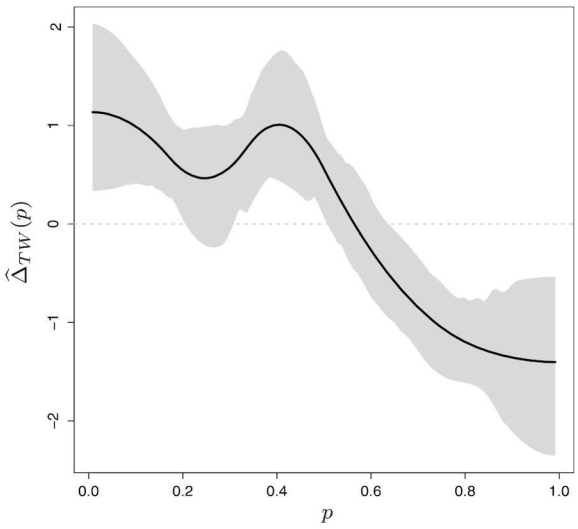

Figure 7: Estimated tailweight difference $\Delta_{TW}(p) = TW_1(p) - TW_2(p)$ between cases and controls, as defined in (22). The solid line represents the estimated pointwise posterior means, while the shaded area shows the corresponding 95% credible intervals.

6 Discussion

In this paper we have introduced a new parameter, the GQTE, for assessing the effect of a binary covariate on a response and a novel methodology to estimate it. The GQTE generalizes the most common approaches available in the literature, that is, the well-known average treatment effect (ATE) and the quantile treatment effect (QTE), since it allows to evaluate the effect of a treatment on any arbitrary characteristic of the outcome's distributions under the two treatment conditions.

To estimate the GQTE we have proposed a Bayesian procedure, where we assume that a monotone transformation of the quantile ratio is modeled as a smooth function of the percentiles. This assumption allows to increase efficiency by borrowing information across the two groups. The idea of quantile ratio smoothing has first been introduced by Dominici et al. (2005). In the present work we extended that proposal in several ways: 1) we let the link between the quantile ratio and the percentiles be general and application-specific, allowing to take into account the tail heaviness of the distributions involved in the analysis; 2) we derive a closed form expression for the model likelihood; 3) our methodology is not limited to the mean difference between the treated and the controls, but provides a comprehensive assessment of the treatment effect; 4) finally, we embed the whole estimation process within a Bayesian framework allowing to make inference on the GQTE $\Delta_g(p)$ for any choice of the function $g(\cdot)$, and for both symmetric and highly skewed outcomes.

The GQTE is a marginal measure in the sense that it provides an estimate of the treatment effect over an entire population. In the econometrics literature this kind of

approach is usually termed the *unconditional* QTE (Firpo, 2007; Frölich and Melly, 2008) in contrast with the *conditional* QTE, where the treatment effect is determined separately for different combinations of a set of covariates (Koenker and Bassett, 1978; Koenker, 2005; Angrist and Pischke, 2009). The inclusion of covariates can improve the efficiency of an estimator even when the primary goal of the analysis is a marginal effect. Accordingly, methods have been proposed to extract marginal quantiles from estimates of conditional quantiles (Machado and Mata, 2005; Frölich and Melly, 2008). A challenge in extending our approach along these lines is the lack of an "iterated expectation" result[1] for the quantiles (see for example Angrist and Pischke, 2009, Chapter 7).

To further clarify our goals, we want to stress that in this paper no particular emphasis has been placed on the causality issues that naturally comes into play when the objective is the estimation of a treatment effect (see for example Rosenbaum, 2002, 2010; Rubin, 2006; Angrist and Pischke, 2009). More precisely, our intent here is solely to provide a general measure of the effect of a binary treatment on a response variable, together with a flexible approach to estimate it.

We compared the performance of our estimation approach with other highly flexible methods in a simulation study for the mean difference between two populations. Our study revealed that the GQTE performs generally better than the other competing estimators at least in estimating the mean difference.

We have applied our methodology to the NMES data set to assess the effect of being affected by smoking attributable diseases on the single hospitalization medical costs distribution. We have found that having these diseases increases the average medical bill amount as well as its variability in the population, while it reduces the probability of incurring higher bills.

Our approach can be extended in various directions. The most promising research question we can see involves taking into account individual level characteristics in measuring the effect of a treatment. In our context, this would involve the estimation of a conditional version of $\Delta_g(p)$, something like $\Delta_g(p|\boldsymbol{x}) = g\left(Q_1(p|\boldsymbol{x})\right) - g\left(Q_2(p|\boldsymbol{x})\right)$. The clear advantage of including covariates would be an increase in the efficiency of the estimates (Frölich and Melly, 2008). To control for systematic differences in covariates between two populations, a common strategy is to group units into subclasses based on covariate values, for example using propensity score matching, and then to apply our method within strata of propensity scores (Rosenbaum, 2002, 2010), as implemented for example in Dominici and Zeger (2005).

Currently we are considering only a binary treatment effect, so another important line of research is the extension of the methods to categorical ordinal and to continuous treatments.

A further direction for future research concerns the choice of the number of degrees of freedom λ. In this paper we adopted the simple approach of choosing λ by minimizing an empirical version of the L_1 distance between the Y_2 density estimate and its kernel density estimate (see Subsection 3.7). More structured solutions can obviously be con-

[1]While for a standard linear model, in fact, the assumption $E(Y_i|X_i) = X_i'\boldsymbol{\beta}$ does imply $E(Y_i) = E(X_i)'\boldsymbol{\beta}$, the same conclusion doesn't hold for the conditional quantiles.

sidered. A natural extension would allow λ to be a random quantity to be estimated together with all the other parameters using a trans-dimensional MCMC approach, like for example the reversible jump algorithm (Green, 1995). While this solution would allow to take into account also the uncertainty connected to the a priori ignorance about the λ value, the consequence would be a dramatic increase in the computational workload of the estimation algorithm.

Acknowledgements

The research of Dominici was supported by Award Number R01ES012054 (Statistical Methods for Population Health Research on Chemical Mixtures) from NIH/NIEHS, Award Numbers R83622 (Statistical Models for Estimating the Health Impact of Air Quality Regulations) and RD83241701 (Estimation of the Risks to Human Health of PM and PM Components) from EPA, Award Number 4909-RFA11-1/12-3 (Causal Inference Methods for Estimating Long Term Health Effects of Air Quality Regulations) from HEI and Award Number K18 HS021991 (A Translational Framework for Methodological Rigor to Improve Patient Centered Outcomes in End of Life Cancer Research) from AHRQ. The content is solely the responsibility of the authors and does not necessarily represent the official views of the above Institutions.

Appendix 1: Additional GQTE Examples

Together with the cases presented in Section 2, many other less conventional measures of the difference between two distributions can be obtained by properly choosing the $g(\cdot)$ function in the GQTE definition. For example, by choosing $g(x) = \int x^r \, dp$ we obtain the difference between the population r-th moments

$$\Delta_{\mu^r} = \int_0^1 Q_1(p)^r \, dp - \int_0^1 Q_2(p)^r \, dp. \tag{19}$$

Using the fact that for a random variable Y with expected value μ, variance σ^2 and quantile function $Q(p)$ it holds that (see Gilchrist, 2000 or Shorack, 2000)

$$\sigma^2 = \int_0^1 [Q(p) - \mu]^2 \, dp = \int_0^1 Q(p)^2 \, dp - \mu^2,$$

by suitably choosing the $g(\cdot)$ function, we recover the difference between the two population variances as

$$\begin{aligned}
\Delta_{\sigma^2} &= \left[\int_0^1 Q_1(p)^2 \, dp - \left(\int_0^1 Q_1(p) \, dp\right)^2\right] - \left[\int_0^1 Q_2(p)^2 \, dp - \left(\int_0^1 Q_2(p) \, dp\right)^2\right] \\
&= \left[\int_0^1 Q_1(p)^2 \, dp - \int_0^1 Q_2(p)^2 \, dp\right] - \left[\left(\int_0^1 Q_1(p) \, dp\right)^2 - \left(\int_0^1 Q_2(p) \, dp\right)^2\right] \\
&= \Delta_{\mu^2} - \left(\mu_1^2 - \mu_2^2\right), \qquad (20)
\end{aligned}$$

However, the cases encompassed by the GQTE include many other quantile-based indexes that are less frequently used in the literature, like the inter-p-range $ipr(p) = Q(1-p) - Q(p)$, or the skewness-ratio $sr(p) = [Q(1-p) - Q(0.5)]/[Q(0.5) - Q(p)]$, $0 < p < 1$, which provide robust measures of the scale and shape of a distribution (for a list of these indexes see Gilchrist, 2000; Shorack, 2000; Parzen, 2004; Wang and Serfling, 2005; Brys et al., 2006). A quantity of particular interest to economists is the difference between inter-decile ratios, defined as

$$\frac{Q_1(0.9)}{Q_1(0.1)} - \frac{Q_2(0.9)}{Q_2(0.1)},$$

which is commonly used to measure the inequality in a population (see Frölich and Melly, 2008). The previous quantity can be easily generalized as follows

$$\Delta_{IR}(p) = \frac{Q_1(1-p)}{Q_1(p)} - \frac{Q_2(1-p)}{Q_2(p)}, \tag{21}$$

for any $0 < p < 0.5$. Notice that all these indexes are obtainable from the general definition (3) by properly choosing the function $g(\cdot)$.

As a last example, we consider a further GQTE special case that is based on the so called *tailweight function* defined as

$$TW(p) = \frac{q(p)}{Q(p)} \equiv \frac{d}{dp} \log Q(p)\,, \qquad 0 < p < 1\,,$$

which is used to quantify the probability allocated in the tails of a distribution. One can compute the difference between the tailweight functions for two populations by choosing the logarithmic derivative of the quantile function as the $g(\cdot)$ functional in (3), that is

$$\begin{aligned} \Delta_{TW}(p) &= TW_1(p) - TW_2(p) \\ &= \frac{d}{dp} \log Q_1(p) - \frac{d}{dp} \log Q_2(p) \\ &= \frac{d}{dp} \left[\log \left(\frac{Q_1(p)}{Q_2(p)} \right) \right]. \end{aligned} \tag{22}$$

If $\Delta_{TW}(p) \geq 0$, we can conclude that the treatment is causing a thickening of the Y_1 distribution tails as compared to those of Y_2 if $\Delta_{TW}(p) \geq 0$. Finally, note that, thanks to the equivariance property of the quantiles, (22) can be written also as

$$\begin{aligned} \Delta_{TW}(p) &= \frac{d}{dp} \log Q_1(p) - \frac{d}{dp} \log Q_2(p) \\ &= \frac{d}{dp} Q_{1,\log}(p) - \frac{d}{dp} Q_{2,\log}(p) \\ &= \frac{d}{dp} \left[Q_{1,\log}(p) - Q_{2,\log}(p) \right] \\ &= \frac{d}{dp} \Delta_{\log}(p)\,, \end{aligned} \tag{23}$$

where $Q_{\ell,\log}$, $\ell = 1, 2$, indicates the quantile of the log-transformed data and $\Delta_{\log}(p)$ denotes the parameter (4) calculated on the quantiles of the log-transformed data.

Appendix 2: Proof of Theorem 1 and Corollaries

Proof of Theorem 1. Differentiate (7) with respect to p to get

$$q_1(p) = q_2(p)\, h^{-1}\left[X(p,\lambda)\,\boldsymbol{\beta}\right] + X'(p,\lambda)\,\boldsymbol{\beta}\, Q_2(p)\left\{\frac{d}{d\left(X(p,\lambda)\,\boldsymbol{\beta}\right)} h^{-1}\left[X(p,\lambda)\,\boldsymbol{\beta}\right]\right\}, \quad (24)$$

where $q_\ell(p) = dQ_\ell(p)/dp$ denotes the so called *quantile density function* for the population $\ell = \{1, 2\}$, while $X'(p,\lambda)$ corresponds to the derivative of $X(p,\lambda)$, $0 < p < 1$ (properly resized because a constant is normally included in the design matrix $X(p,\lambda)$).

Apply now to both $q_1(p)$ and $q_2(p)$ the following relationship between the quantile density and density quantile functions (see for example Gilchrist, 2000; Parzen, 1979, 2004)

$$f(Q(p))\, q(p) = 1\,, \quad (25)$$

to get the expression

$$\frac{1}{f_1(Q_1(p)|\boldsymbol{\eta})} = \frac{1}{f_2(Q_2(p)|\boldsymbol{\beta},\boldsymbol{\eta})}\; h^{-1}\left[X(p,\lambda)\,\boldsymbol{\beta}\right] + X'(p,\lambda)\,\boldsymbol{\beta}\, Q_2(p)\left\{\frac{d}{d\left(X(p,\lambda)\,\boldsymbol{\beta}\right)} h^{-1}\left[X(p,\lambda)\,\boldsymbol{\beta}\right]\right\}, \quad (26)$$

and hence

$$f_2(Q_2(p)|\boldsymbol{\beta},\boldsymbol{\eta}) = \frac{f_1(Q_1(p)|\boldsymbol{\eta}) h^{-1}\left[X(p,\lambda)\,\boldsymbol{\beta}\right]}{1 - f_1(Q_1(p)|\boldsymbol{\eta})\, X'(p,\lambda)\,\boldsymbol{\beta}\, Q_2(p)\left\{\frac{d}{d\left(X(p,\lambda)\,\boldsymbol{\beta}\right)} h^{-1}\left[X(p,\lambda)\,\boldsymbol{\beta}\right]\right\}}\,. \quad (27)$$

Finally, substituting (7) in place of $Q_1(p)$ proves the main statement.

Moreover, $f_2(Q_2(p)|\boldsymbol{\beta},\boldsymbol{\eta})$ is a proper density function because:

- $f_2(Q_2(p)|\boldsymbol{\beta},\boldsymbol{\eta}) \geq 0$, for any $0 < p < 1$; since $f_1\left(Q_1(p)|\boldsymbol{\eta}\right) \geq 0$ and $h^{-1}\left[X(p,\lambda)\,\boldsymbol{\beta}\right] > 0$ because, as assumed in (6), it is the ratio of two positive quantile functions, this fact can be proved by showing that the denominator of (27) is nonnegative which is ensured by the constraint (10).
- $\int_0^1 f_2(Q_2(p)|\boldsymbol{\beta},\boldsymbol{\eta}) q_2(p) dp = 1$, which is true because $f_2(Q_2(p)|\boldsymbol{\beta},\boldsymbol{\eta}) q_2(p) = 1$ by construction. □

A couple of immediate consequences of Theorem 1 regard two cases that occur frequently in practice. We provide the details about these situations in the next two corollaries.

Corollary 1. *Let the same assumptions of Theorem 1 hold. Suppose additionally that* $h(x) = x$. *If for every* $0 < p < 1$ *the vector* $\boldsymbol{\beta}$ *satisfies the constraint*

$$\frac{X'(p,\lambda)\,\boldsymbol{\beta}}{X(p,\lambda)\,\boldsymbol{\beta}} \leq \frac{1}{f_1\left(Q_1(p)\right) Q_1(p)}\,, \quad (28)$$

then the density quantile function $f_2(Q_2(p)|\boldsymbol{\beta},\boldsymbol{\eta})$ *for* Y_2 *is*

$$f_2(Q_2(p)|\boldsymbol{\beta},\boldsymbol{\eta}) = \frac{f_1\left(Q_2(p)\, X(p,\lambda)\,\boldsymbol{\beta}|\boldsymbol{\eta}\right) X(p,\lambda)\,\boldsymbol{\beta}}{1 - f_1\left(Q_2(p)\, X(p,\lambda)\,\boldsymbol{\beta}|\boldsymbol{\eta}\right) X'(p,\lambda)\,\boldsymbol{\beta}\, Q_2(p)}\,. \tag{29}$$

Corollary 2. *Let the same assumptions of Theorem 1 hold. Suppose additionally that* $h(x) = \log(x)$. *If for every* $0 < p < 1$ *the vector* $\boldsymbol{\beta}$ *satisfies the constraint*

$$X'(p,\lambda)\,\boldsymbol{\beta} \leq \frac{1}{f_1\left(Q_1(p)\right) Q_1(p)}\,, \tag{30}$$

then the density quantile function $f_2(Q_2(p)|\boldsymbol{\beta},\boldsymbol{\eta})$ *for* Y_2 *is given by*

$$f_2(Q_2(p)|\boldsymbol{\beta},\boldsymbol{\eta}) = \frac{f_1\left(Q_2(p)\, e^{\,X(p,\lambda)\,\boldsymbol{\beta}}|\boldsymbol{\eta}\right)}{e^{-\,X(p,\lambda)\,\boldsymbol{\beta}} - f_1\left(Q_2(p)\, e^{\,X(p,\lambda)\,\boldsymbol{\beta}}|\boldsymbol{\eta}\right) X'(p,\lambda)\,\boldsymbol{\beta}\, Q_2(p)}\,. \tag{31}$$

Note that in these two situations, the general constraint (10) reduces to a linear constraint on $\boldsymbol{\beta}$.

Appendix 3: Proofs of the Special Cases

Case 1: Y_1 *is Uniform and* $X(p,\lambda=0)=1$. Here $Y_1|\theta_1 \sim \mathcal{U}[0,\theta_1]$ and $h(x)=x$. In this case $Q_1(p)/Q_2(p) = \beta_0$. Hence the density, distribution and quantile functions of Y_1 are respectively

$$\begin{aligned} f_1(y_1|\theta_1) &= \frac{1}{\theta_1}\,\mathbb{I}_{[0,\,\theta_1]}\{y_1\} \\ F_1(y_1|\theta_1) &= \frac{y_1}{\theta_1} \\ Q_1(p|\theta_1) &= \theta_1 p\,, \qquad\qquad 0 < p < 1\,. \end{aligned}$$

From (29) it follows that

$$\begin{aligned} f_2(Q_2(p)|\theta_1,\beta_0) &= \frac{1}{\theta_1}\,\mathbb{I}_{[0,\,\theta_1]}\{Q_2(p)\,\beta_0\}\,\beta_0 \\ &= \frac{\beta_0}{\theta_1}\,\mathbb{I}_{[0,\,\theta_1/\beta_0]}\{Q_2(p)\}\,, \end{aligned}$$

which is the density quantile function of a $\mathcal{U}[0,\theta_2]$ random variable with $\theta_2 = \theta_1/\beta_0$.

Case 2: Y_1 *is Log-normal and* $X(p,\lambda=1) = [1, \Phi^{-1}(p)]$. Assume $Y_1|\mu_1,\sigma_1^2 \sim \mathcal{L}n(\mu_1,\sigma_1^2)$ and $h(x)=\log(x)$. In this case $\log\{Q_1(p)/Q_2(p)\} = \beta_0 + \beta_1\,\Phi^{-1}(p)$, where $\Phi^{-1}(p)$ is the quantile function of a standard normal random variable. The density, distribution and quantile functions of Y_1 are given by

$$f_1(y_1|\mu_1,\sigma_1^2) = \frac{1}{y_1\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{(\log y_1 - \mu_1)^2}{2\sigma_1^2}\right\}$$

$$
\begin{aligned}
F_1(y_1|\mu_1,\sigma_1^2) &= \Phi\left(\frac{\log y_1-\mu_1}{\sigma_1}\right)\\
Q_1(p|\mu_1,\sigma_1^2) &= \exp\left\{\,\mu_1+\sigma_1\Phi^{-1}(p)\right\}\,, \qquad 0<p<1\,.
\end{aligned}
$$

Then by (31) it follows

$$
\begin{aligned}
f_2(Q_2(p)|\mu_1,\sigma_1^2,\beta_0,\beta_1) &= \frac{\frac{1}{\sqrt{2\pi}\sigma_1}\,\frac{\exp\left\{-\frac{\left(\mu_1+\sigma_1\Phi^{-1}(p)-\mu_1\right)^2}{2\sigma_1^2}\right\}}{\exp\left\{\,\mu_1+\sigma_1\Phi^{-1}(p)\right\}}}{\exp\left\{-\beta_0-\beta_1\Phi^{-1}(p)\right\}\left[1-\frac{1}{\sqrt{2\pi}\sigma_1}\,\frac{\exp\left\{-\frac{\left(\mu_1+\sigma_1\Phi^{-1}(p)-\mu_1\right)^2}{2\sigma_1^2}\right\}}{\exp\left\{\,\mu_1+\sigma_1\Phi^{-1}(p)\right\}}\right.}\\
&\qquad \left.\times\beta_1\frac{1}{\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{\left[\Phi^{-1}(p)\right]^2}{2}\right\}}\,\exp\left\{\,\mu_1+\sigma_1\Phi^{-1}(p)\right\}\right]}\\
&= \frac{1}{\sqrt{2\pi}(\sigma_1-\beta_1)}\,\frac{\exp\left\{-\frac{\left[\Phi^{-1}(p)\right]^2}{2}\right\}}{\exp\left\{\,(\mu_1-\beta_0)+(\sigma_1-\beta_1)\Phi^{-1}(p)\right\}}\\
&= \frac{1}{\sqrt{2\pi}(\sigma_1-\beta_1)}\,\frac{\exp\left\{-\frac{\left[(\mu_1-\beta_0)+(\sigma_1-\beta_1)\Phi^{-1}(p)-(\mu_1-\beta_0)\right]^2}{2(\sigma_1-\beta_1)^2}\right\}}{\exp\left\{\,(\mu_1-\beta_0)+(\sigma_1-\beta_1)\Phi^{-1}(p)\right\}}\\
&= \frac{1}{Q_2(p)\sqrt{2\pi}(\sigma_1-\beta_1)}\,\exp\left\{-\frac{\left[\log Q_2(p)-(\mu_1-\beta_0)\right]^2}{2(\sigma_1-\beta_1)^2}\right\}\,,
\end{aligned}
$$

which is the density quantile function of a $\mathcal{L}n(\mu_2,\sigma_2^2)$ random variable with $\mu_2=(\mu_1-\beta_0)$ and $\sigma_2=(\sigma_1-\beta_1)$.

Case 3: Y_1 is Pareto and $X(p,\lambda=1)=[1,\log(1-p)]$. Now $Y_1|a_1,b_1\sim\mathcal{P}a(a_1,b_1)$ and $h(x)=\log(x)$. In this case $\log\{Q_1(p)/Q_2(p)\}=\beta_0+\beta_1\log(1-p)$. The density, distribution and quantile functions of Y_1 are given by

$$
\begin{aligned}
f_1(y_1|a_1,b_1) &= a_1 b_1^{a_1} y_1^{-(a_1+1)}\\
F_1(y_1|a_1,b_1) &= 1-b_1^{a_1}y_1^{-a_1}\\
Q_1(p|a_1,b_1) &= b_1(1-p)^{-\frac{1}{a_1}}\,, \qquad 0<p<1\,.
\end{aligned}
$$

Then (31) implies

$$
\begin{aligned}
f_2(Q_2(p)|a_1,b_1,\beta_0,\beta_1) &= \frac{a_1 b_1^{a_1}\left[b_1(1-p)^{-\frac{1}{a_1}}\right]^{-(a_1+1)}}{\exp\{-\beta_0-\beta_1\log(1-p)\}\left\{1+a_1 b_1^{a_1}\left[b_1(1-p)^{-\frac{1}{a_1}}\right]^{-(a_1+1)}\frac{\beta_1}{1-p}b_1(1-p)^{-\frac{1}{a_1}}\right\}}\\
&= \frac{a_1}{a_1\beta_1+1}\left(b_1e^{-\beta_0}\right)^{-1}(1-p)^{\frac{a_1\beta_1+1}{a_1}+1}\\
&= \frac{a_1}{a_1\beta_1+1}\left(b_1e^{-\beta_0}\right)^{\frac{a_1}{a_1\beta_1+1}}Q_2(p)^{-\left(\frac{a_1}{a_1\beta_1+1}+1\right)}\,,
\end{aligned}
$$

which is the density quantile function of a $\mathcal{P}a(a_2, b_2)$ random variable with $a_2 = \frac{a_1}{a_1\beta_1+1}$ and $b_2 = b_1 e^{-\beta_0}$.

Appendix 4: Details About the Estimation of Other Cases

One can estimate the impact of a binary treatment on the r-th moments, denoted as Δ_{μ^r} in (19), by computing

$$\widehat{\Delta}_{\mu^r}^{(m)} = \frac{1}{n_2}\sum_{i=1}^{n_2}\left\{y_{2(i)}h^{-1}\left[X\left(p_{2i},\lambda\right)\widehat{\boldsymbol{\beta}}^{(m)}\right]\right\}^r - \frac{1}{n_1}\sum_{i=1}^{n_1}\left[y_{1(i)}\left\{h^{-1}\left[X\left(p_{1i},\lambda\right)\widehat{\boldsymbol{\beta}}^{(m)}\right]\right\}^{-1}\right]^r. \quad (32)$$

The last expression allows to estimate the treatment effect on the population variances, defined in (20), which is given by

$$\widehat{\Delta}_{\sigma^2}^{(m)} = \widehat{\Delta}_{\mu^2}^{(m)} - \left\{\left(\frac{1}{n_2}\sum_{i=1}^{n_2} y_{2(i)}h^{-1}\left[X\left(p_{2i},\lambda\right)\widehat{\boldsymbol{\beta}}^{(m)}\right]\right)^2 - \left(\frac{1}{n_1}\sum_{i=1}^{n_1} y_{1(i)}\left\{h^{-1}\left[X\left(p_{1i},\lambda\right)\widehat{\boldsymbol{\beta}}^{(m)}\right]\right\}^{-1}\right)^2\right\}. \quad (33)$$

As a concluding example, the effect of a binary treatment on the tailweight functions of two distributions, introduced in (22), can be obtained by first computing the posterior draws

$$\widehat{\Delta}_{TW}^{(m)}(p) = \frac{d}{dp}\left\{\log\left(h^{-1}\left[X(p,\lambda)\widehat{\boldsymbol{\beta}}^{(m)}\right]\right)\right\}, \quad (34)$$

and then by applying (15). When $h(x) = \log(x)$, (34) becomes

$$\widehat{\Delta}_{TW}^{(m)}(p) = X'(p,\lambda)\widehat{\boldsymbol{\beta}}^{(m)}, \quad (35)$$

and the estimate of $\Delta_{TW}(p)$ is

$$\begin{aligned}\widehat{\Delta}_{TW}(p) &= \frac{1}{M}\sum_{m=1}^{M} X'(p,\lambda)\widehat{\boldsymbol{\beta}}^{(m)} \\ &= X'(p,\lambda)\left(\frac{1}{M}\sum_{m=1}^{M}\widehat{\boldsymbol{\beta}}^{(m)}\right) \\ &= X'(p,\lambda)\widehat{\boldsymbol{\beta}}, \end{aligned} \quad (36)$$

with $\widehat{\boldsymbol{\beta}} = \frac{1}{M}\sum_{m=1}^{M}\widehat{\boldsymbol{\beta}}^{(m)}$, the posterior mean estimate of $\boldsymbol{\beta}$.

References

Abadie, A., Angrist, J. D., and Imbens, G. (2002). "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings." *Econometrica*, 70: 91–117. 524

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics*. Princeton University Press, Princeton, NJ. 543

Brys, G., Hubert, M., and Struyf, A. (2006). "Robust measures of tail weight." *Computational Statistics & Data Analysis*, 50(3): 733–759. 545

Cameron, C. A. and Trivedi, P. K. (2005). *Microeconometrics*. Cambridge University Press, New York. 541

Carlin, B. P. and Louis, T. A. (2009). *Bayesian methods for data analysis*. Chapman & Hall/CRC, Boca Raton, Third edition. 530

Chen, Y. and Dunson, D. B. (2009). "Nonparametric Bayes conditional distribution modeling with variable selection." *Journal of the American Statistical Association*, 104(488): 1646–1660. 532

Chernozhukov, V. and Hansen, C. (2005). "An IV model of quantile treatment effects." *Econometrica*, 73: 245–261. 524

De Iorio, M., Müller, P., Ronser, G. L., and MacEachern, S. N. (2004). "An ANOVA model for dependent random measures." *Journal of the American Statistical Association*, 99(465): 205–215. 533

Devroye, L. and Lugosi, G. (2001). *Combinatorial methods in density estimation*. Springer. 532

Dominici, F., Cope, L., Naiman, D. Q., and Zeger, S. L. (2005). "Smooth quantile ratio estimation (SQUARE)." *Biometrika*, 92: 543–557. 523, 524, 526, 532, 533, 535, 542

Dominici, F. and Zeger, S. L. (2005). "Smooth quantile ratio estimation with regression: estimating medical expenditures for smoking-attributable diseases." *Biostatistics*, 6: 505–519. 543

Dominici, F., Zeger, S. L., Parmigiani, G., Katz, J., and Christian, P. (2006). "Estimating percentile-specific treatment effects in counterfactual models: a case-study of micronutrient supplementation, birth weight and infant mortality." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55: 261–280. 525

— (2007). "Does the effect of micronutrient supplementation on neonatal survival vary with respect to the percentiles of the birth weight distribution?" *Bayesian Analysis*, 2: 1–30. 525

Firpo, S. (2007). "Efficient semiparametric estimation of quantile treatment effects." *Econometrica*, 75: 259–276. 524, 543

Frölich, M. and Melly, B. (2008). "Unconditional quantile treatment effects under endogeneity." Technical Report 3288, Institute for the Study of Labor (IZA), P.O. Box 7240, 53072 Bonn, Germany. 524, 525, 543, 545

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. Chapman & Hall/CRC, Boca Raton, Third edition. 531

Gilchrist, W. G. (2000). *Statistical modelling with quantile functions*. Chapman & Hall/CRC, New York. 528, 544, 545, 546

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (eds.) (1996). *Markov Chain Monte Carlo in practice*. Chapman & Hall/CRC, New York. 530

Green, P. J. (1995). "Reversible jump MCMC computation and Bayesian model determination." *Biometrika*, 82(4): 711–732. 544

Koenker, R. (2005). *Quantile regression*. Cambridge University Press, New York. 543

Koenker, R. and Bassett, G. S. (1978). "Regression quantiles." *Econometrica*, 46: 33–50. 543

MacEachern, S. N. (1999). "Dependent Nonparametric Processes." In *ASA Proceedings of the Section on Bayesian Statistical Science*. Alexandria, VA: American Statistical Association. 533

Machado, J. and Mata, J. (2005). "Counterfactual decompositions of changes in wage distributions using quantile regression." *Journal of Applied Econometrics*, 20: 445–465. 543

Mullahy, J. (1998). "Much ado about two: reconsidering retransformation and the two-part model in health econometrics." *Journal of Health Economics*, 17: 247–281. 541

National Center For Health Services Research (1987). *National Medical Expenditure Survey*. National Center for Health Services Research and Health Technology Assessment. 535

O'Hagan, G. and Forster, J. (2004). *Bayesian inference*. Arnold, London, Second edition. 530

Parzen, E. (1979). "Nonparametric statistical data modeling." *Journal of the American Statistical Association*, 74: 105–121. 528, 546

— (2004). "Quantile probability and statistical data modeling." *Statistical Science*, 19: 652–662. 545, 546

Robert, C. P. and Casella, G. (2004). *Monte Carlo statistical methods*. Springer, New York, Second edition. 530

Rosenbaum, P. (2002). *Observational studies*. Springer, New York. 543

— (2010). *Design of observational studies*. Springer, New York. 543

Rubin, D. B. (2006). *Matched sampling for causal effects*. Cambridge University Press, Cambridge, UK. 543

Sethuraman, J. (1994). "A constructive definition of Dirichlet priors." *Statistica Sinica*, 4: 639–650. 533

Shorack, G. R. (2000). *Probability for statisticians*. Springer, New York. 544, 545

Venturini, S., Dominici, F., and Parmigiani, G. (2008). "Gamma shape mixtures for heavy-tailed distributions." *Annals of Applied Statistics*, 2: 756–776. 536, 538

Wang, J. and Serfling, R. (2005). "Nonparametric multivariate kurtosis and tailweight measures." *Journal of Nonparametric Statistics*, 17: 441–456. 545

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press, Second edition. 525